The Introduction To Artificial Intelligence

Yuni Zeng yunizeng@zstu.edu.cn
2024-2025-1

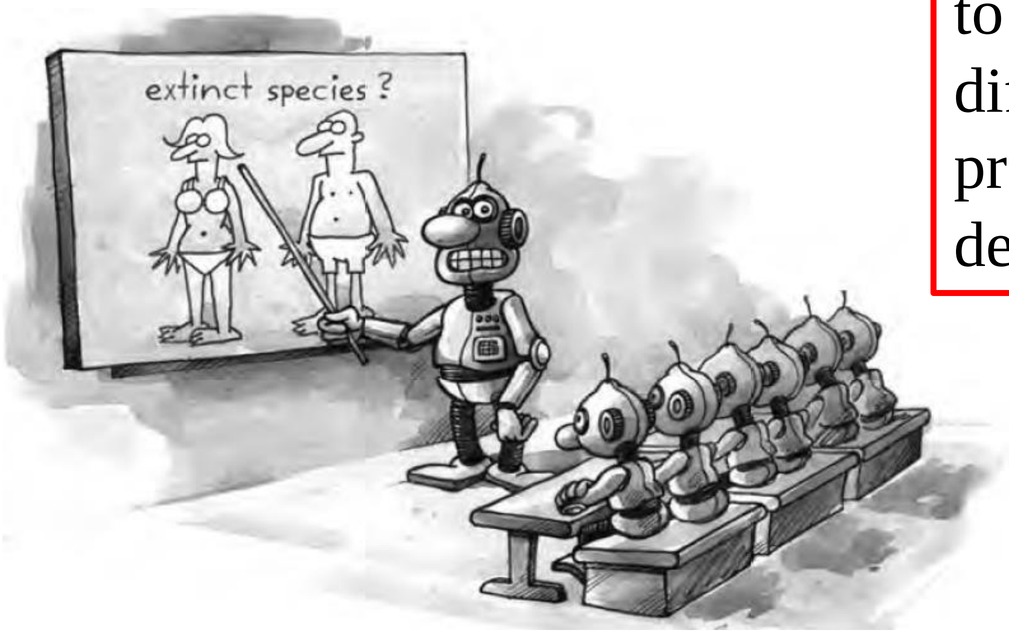
The Introduction to Artificial Intelligence

- Part I Brief Introduction to AI & Different AI tribes
- Part II Knowledge Representation & Reasoning
- Part III AI GAMES and Searching
- Part IV Model Evaluation and Selection
- Part V Machine Learning

What is learning?

- Memorize the words in a vocabulary?
- Learn to complete the addition : x+y?
- How do you learn the addition between two numbers?

Machine learning allows us to tackle tasks that are too difficult to solve with fixed programs written and designed by human beings.



What is Machine Learning?

- Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed.
- A machine learning algorithm is an algorithm that is able to learn from data.
- The development of machine learning algorithms is one of the most important branches of AI.

What is Machine Learning?

A widely quoted and formal definition is "*A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E*"

"如果一个程序在某类任务T中，受性能指标P的度量，其性能能随着经验值E的上升而不断提升，这个程序就能从与任务T和性能指标P相关的经验值E中学习。"

Machine Learning

- *1. Different ML methods*
- 2. Data representation
- 3. Data preprocessing

1. Different ML methods

Supervised learning

Unsupervised learning

Reinforcement learning

1. Different ML methods

Supervised learning

Unsupervised learning

Reinforcement learning

1. Different ML methods

Supervised Learning

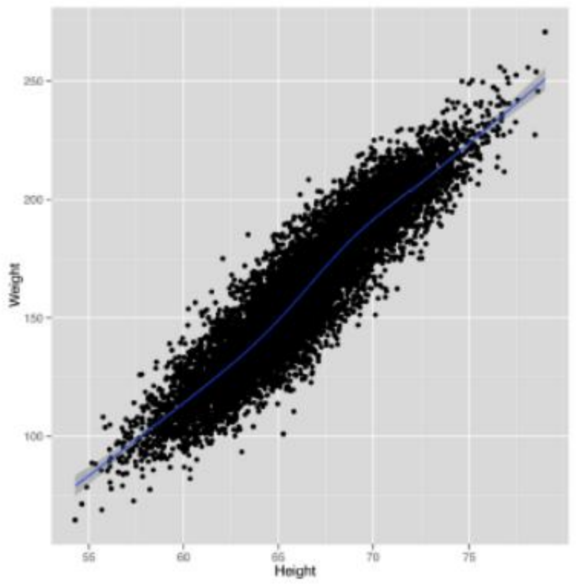
Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs.

- Regression
- Classification

1. Different ML methods

- What is regression?

Regression is to relate input variables to the output variable, to either predict outputs for new inputs and/or to interpret the effect of the input on the output.



Height is correlated with weight.

1. Different ML methods

Two goals of regression

Prediction

wish to predict the output for a new input vector

- e.g. What is the weight of a person who is 170 cm tall?

Interpretation

Understand the effect of inputs on output

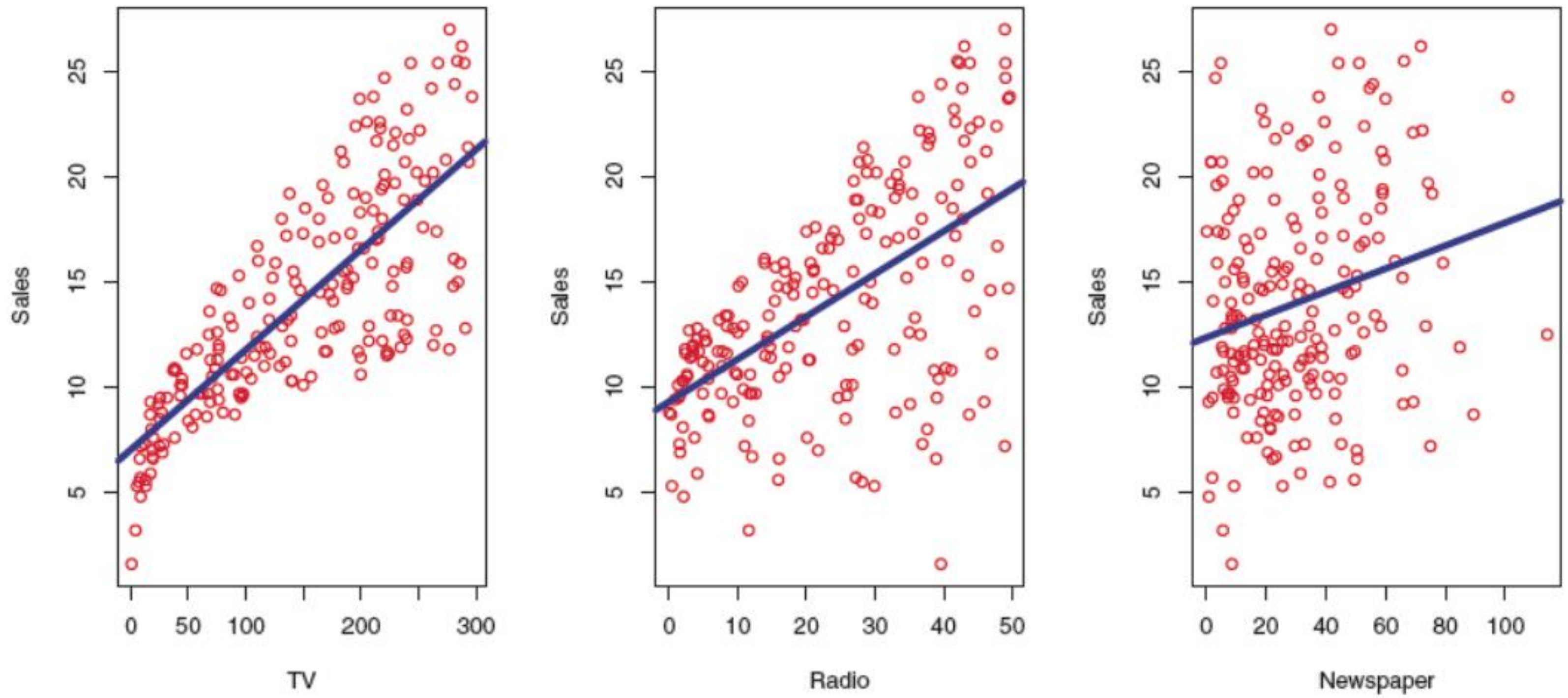
- e.g. Are taller people heavier too?

For both the goals, we need to find a function that approximates the output "well enough" given inputs.

$$\boxed{y_n \approx f(x_n), \text{ for all } n}$$

1. Different ML methods

Regression --- example

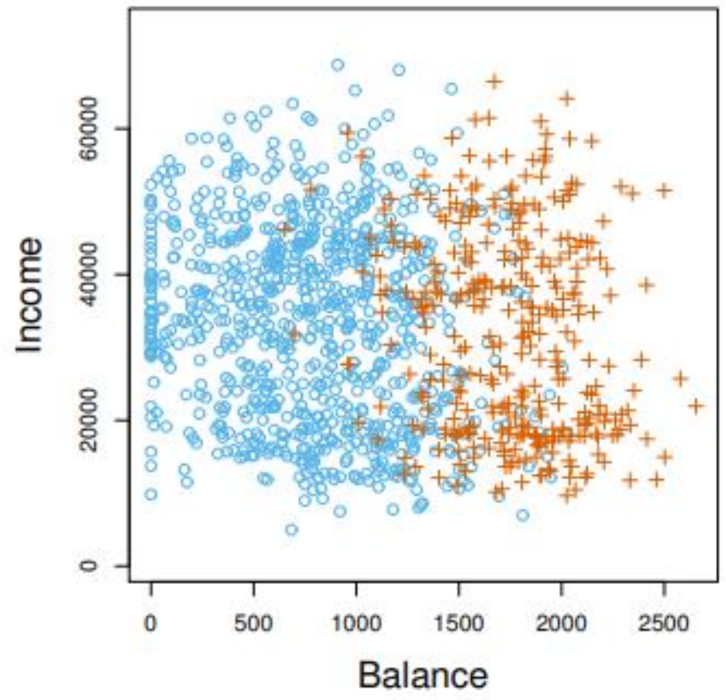


How does advertisement in TV, radio, and newspaper affect sales?

1. Different ML methods

Classification

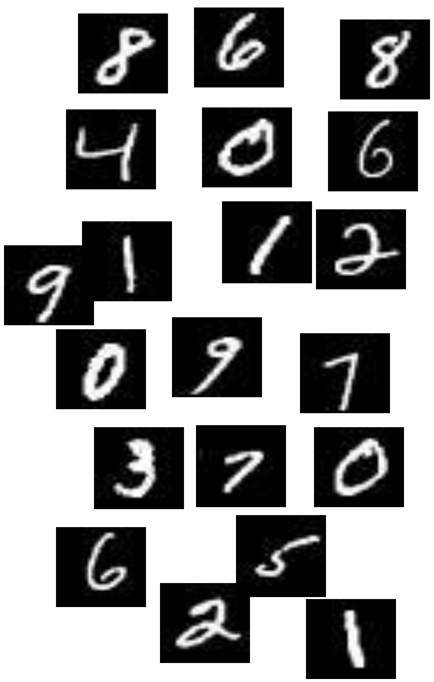
- Classification is same as regression but now y_n is binary or has finite values.
- Examples: object detection, face detection, hand-written digits recognition.



1. Different ML methods

Classification – An example

Training set



My baby, I will show you the digits today.
Let's repeat it again…

 eight.

It's eight.

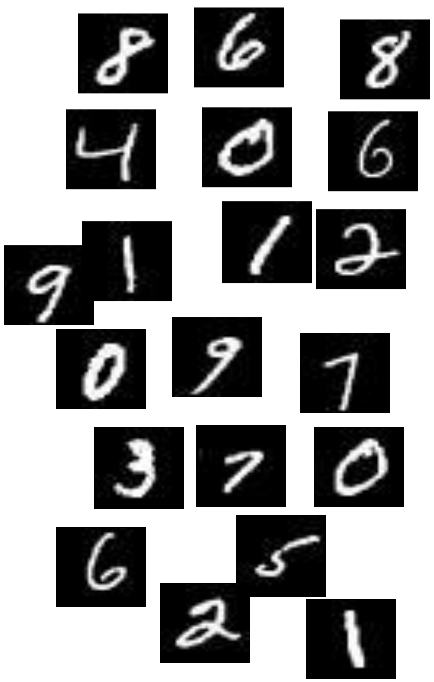
Mother knows the label for each training data.

1. Different ML methods

- Classification – An example

Training now …

Training set



Let me see whether you know these digits.

Training set

Yes! You know all the digits!

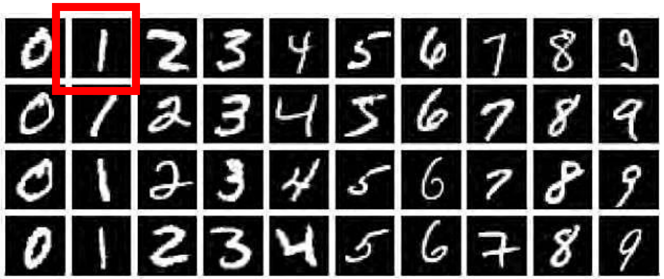
Yes, I know all of them. They are eight, six, eight, four…

Recognition accuracy on training set:100%

1. Different ML methods

Classification – An example

The testing set: independent from the training set.



The central challenge in machine learning is that we must perform well on inputs—not just those on which our model was trained. The ability to perform well on previously unobserved inputs is called generalization.

My baby, I will test you on what you learned.

What's this?

It's one. √

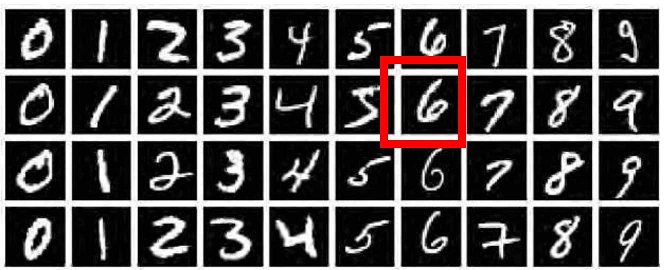
Testing now …

1. Different ML methods

Classification – An example

My baby, I will test you on what you learned.

The testing set: independent from the training set.



The central challenge in machine learning is that we must perform well on inputs—not just those on which our model was trained. The ability to perform well on previously unobserved inputs is called generalization.

What's this?

It's six. √

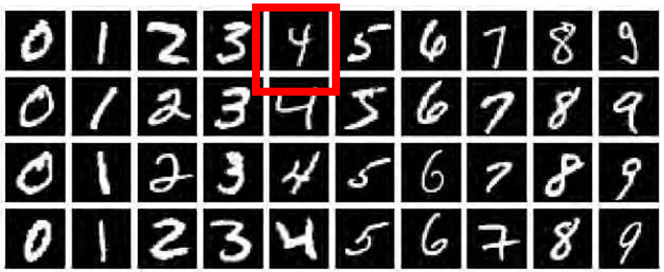
Testing now …

1. Different ML methods

Classification – An example

My baby, I will test you on what you learned.

The testing set: independent from the training set.



The central challenge in machine learning is that we must perform well on inputs—not just those on which our model was trained. The ability to perform well on previously unobserved inputs is called generalization.

It's nine. X

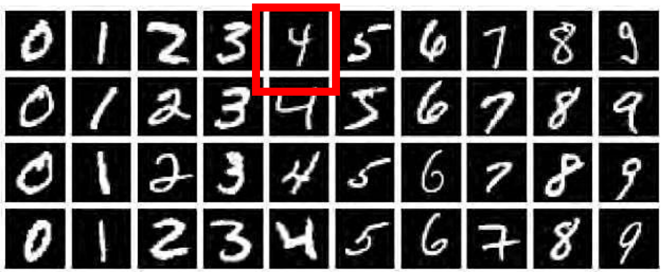
Testing now …

1. Different ML methods

Classification – An example

My baby, I will test you on what you learned.

The testing set: independent from the training set.



The central challenge in machine learning is that we must perform well on inputs—not just those on which our model was trained. The ability to perform well on previously unobserved inputs is called generalization.

Let me see.
You've answered 38 of the 40 digits correctly.
So you scored 95 points (38/40=95%).

Recognition accuracy on testing set :95%

Testing now …

1. Different ML methods

Supervised learning

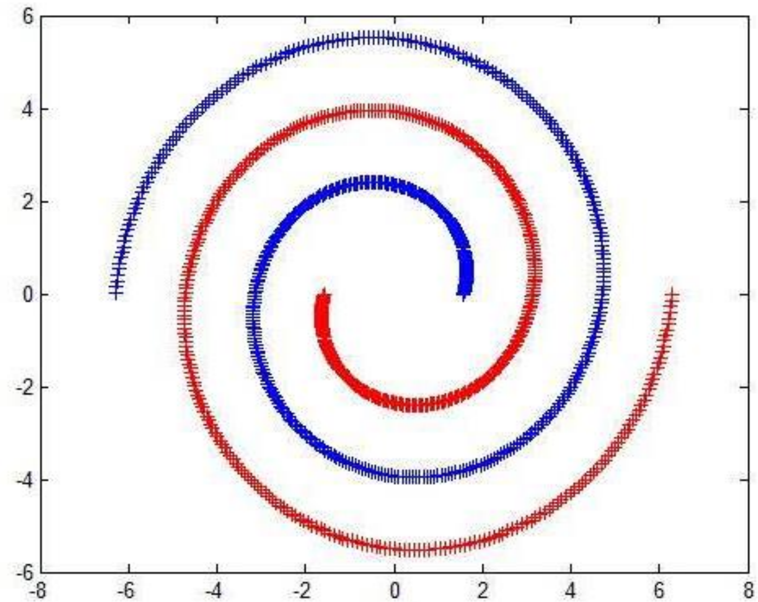
Unsupervised learning

Reinforcement learning

1. Different ML methods

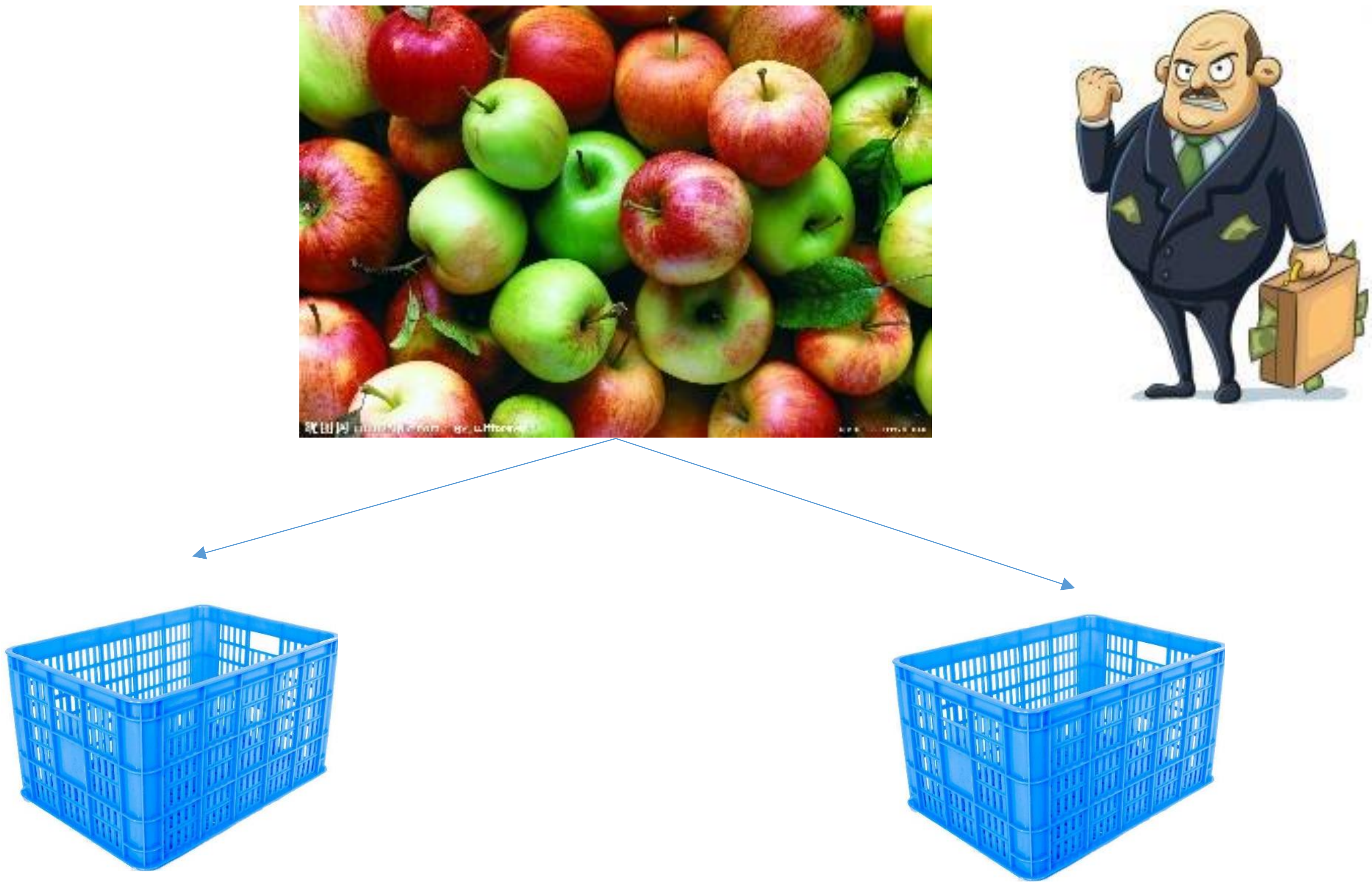
- Clustering

  - **Unsupervised machine learning** is the machine learning task of inferring a function that describes the structure of *"unlabeled" data.*



1. Different ML methods

- Clustering



1. Different ML methods

- Clustering

1. Different ML methods

- Clustering

I give you some examples.

1. Different ML methods

Supervised learning	Semi-supervised learning	Unsupervised learning

- **Semi-supervised learning** is a class of techniques that make use of unlabeled data for training.
- There are typically a small amount of labeled data with a large amount of unlabeled data

1. Different ML methods

Supervised learning

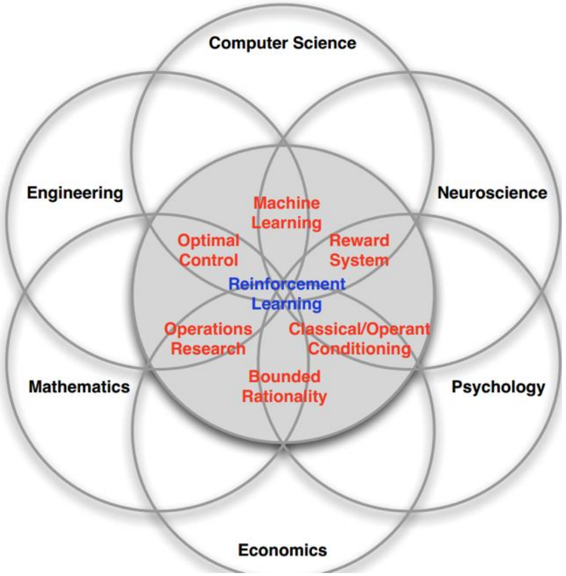
Unsupervised learning

Reinforcement learning

1. Different ML methods

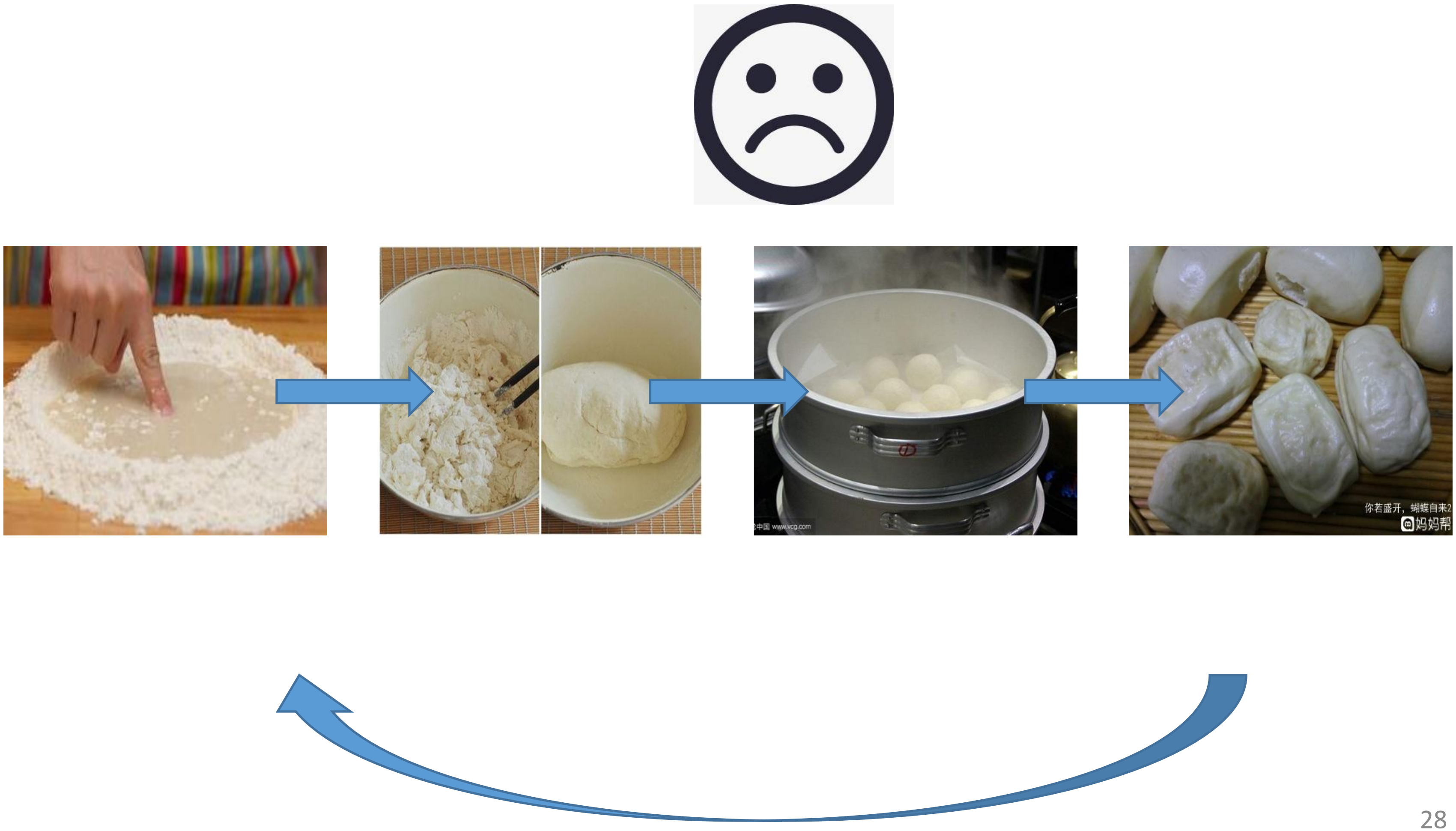
Reinforcement Learning

- "AI=RL" by David Silver
- Agent-oriented learning—learning by interacting with an environment to achieve a goal
- Learning by trial and error, with only delayed evaluative feedback (reward)



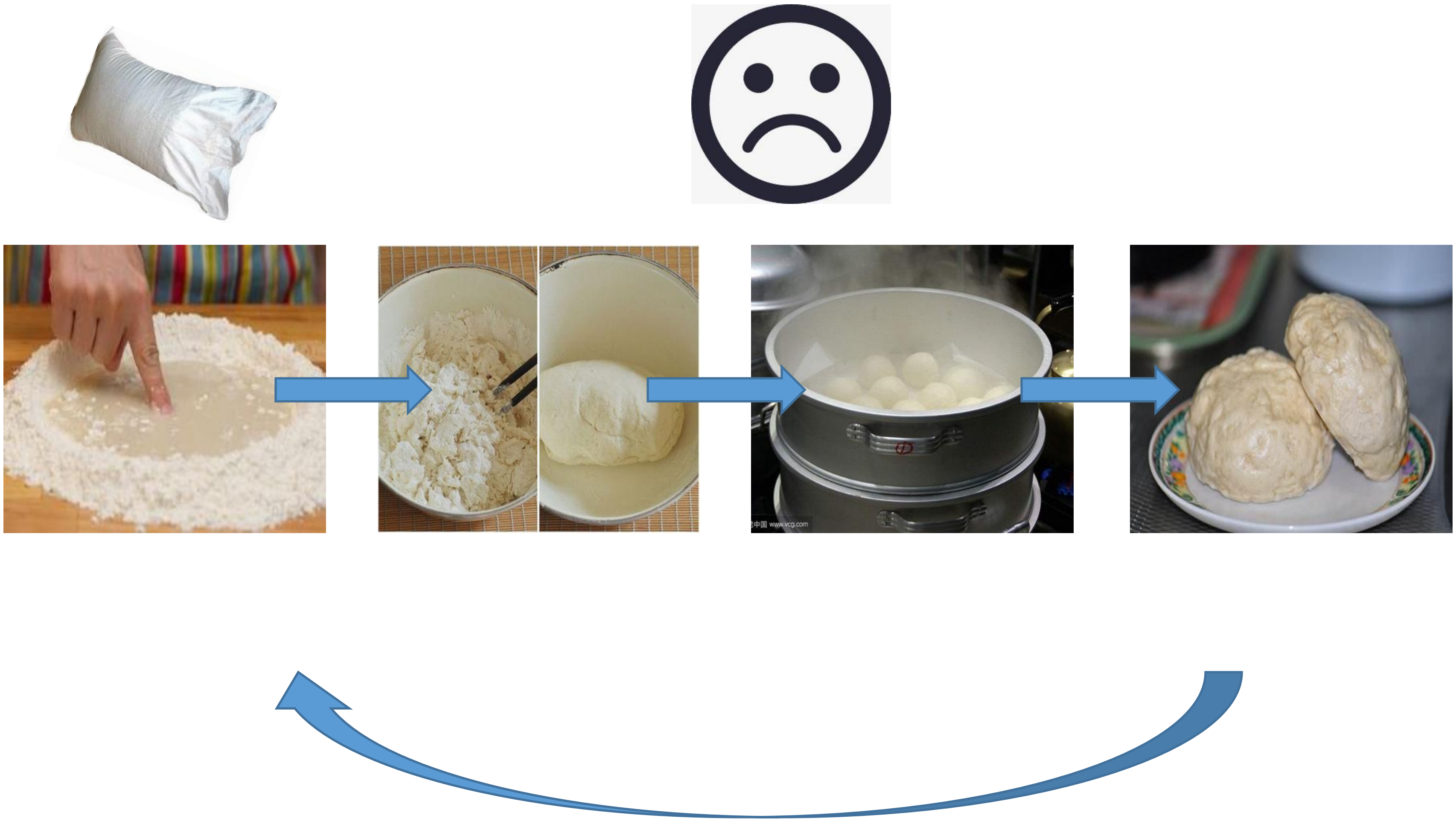
1. Different ML methods

- Reinforcement Learning -- example



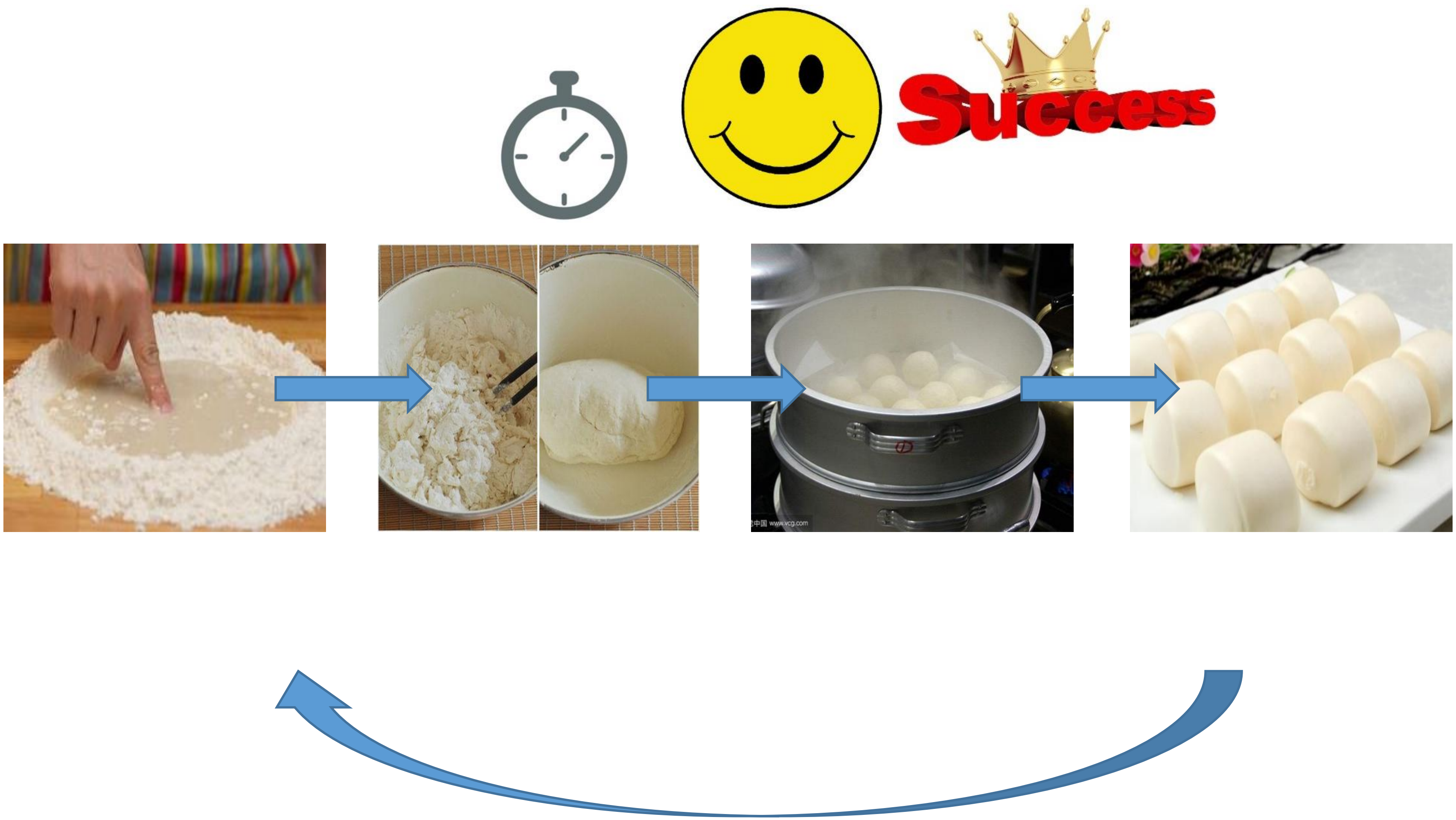
1. Different ML methods

- Reinforcement Learning -- example



1. Different ML methods

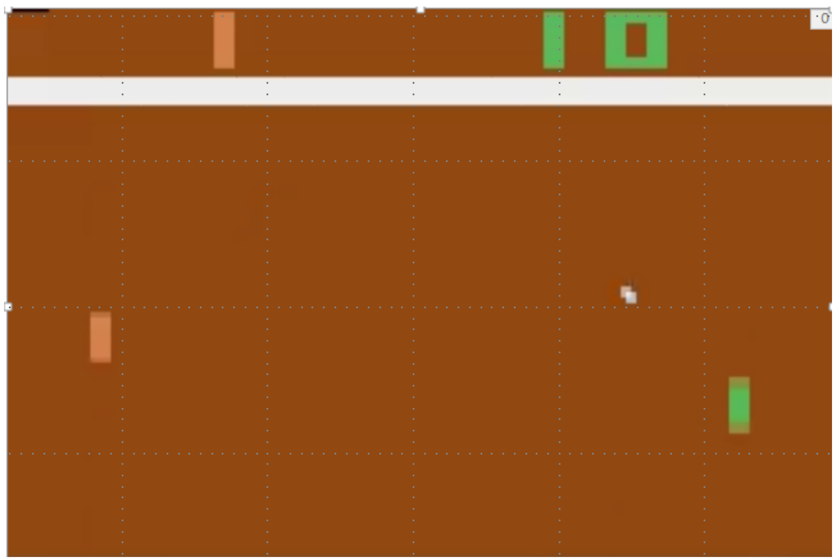
- Reinforcement Learning -- example



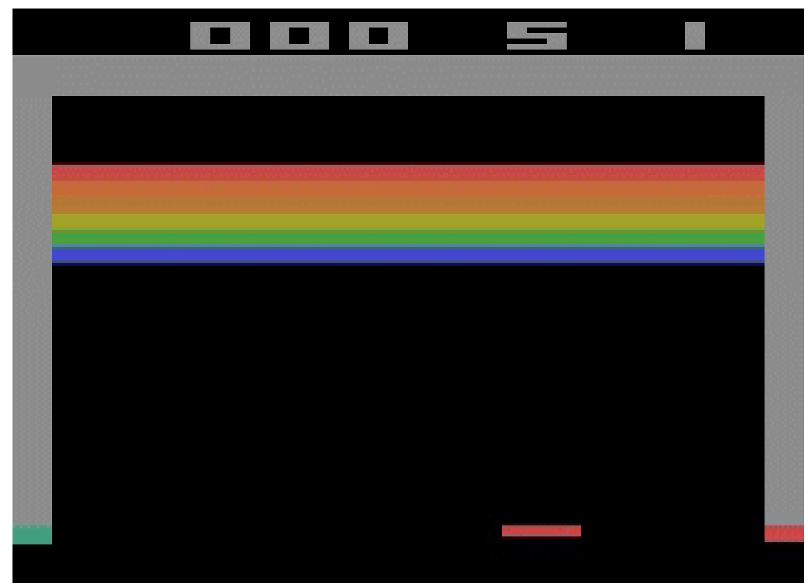
1. Different ML methods

Reinforcement Learning

- **Game Pong**

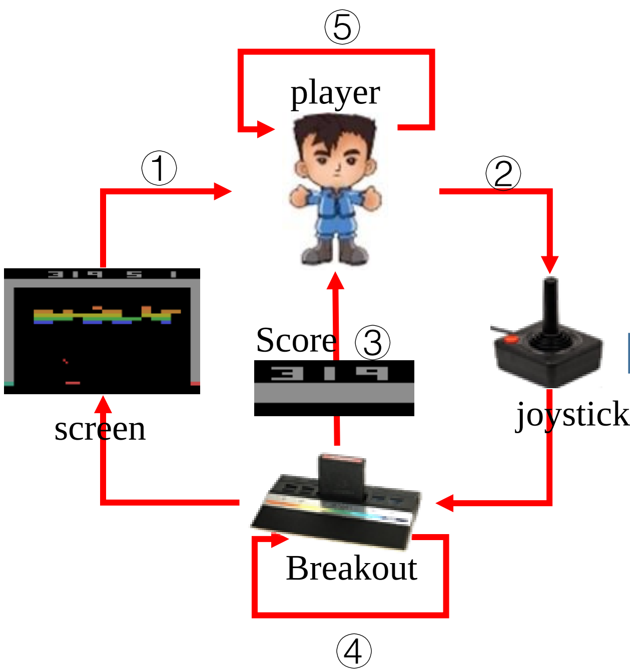


- **Game Breakout**

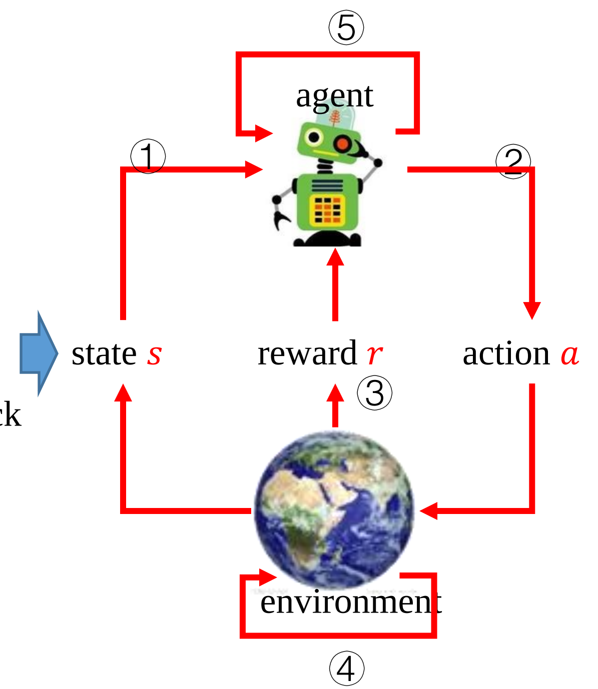


1. Different ML methods

Reinforcement Learning



Game Breakout



Reinforcement Learning

- Rules are unknown
- Learn directly from the interaction

At each time step t:

① Agent receives state $\boldsymbol{s}(\boldsymbol{t})$

② Agent executes an action $\boldsymbol{a}(\boldsymbol{t})$ by his action policy $\boldsymbol{\pi}(s(t))$

③ Environment emits a immediate reward $\boldsymbol{r}(\boldsymbol{t}+\boldsymbol{1})$ to agent

④ Environment changes its state to $\boldsymbol{s}(\boldsymbol{t}+\boldsymbol{1})$

⑤ Agent improves his policy $\boldsymbol{\pi}(\boldsymbol{s})$ according to the reward.

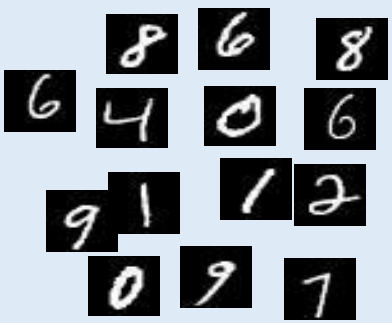
$$\begin{cases} < s, a, r, s' > \\ s \leftarrow s' \end{cases}$$

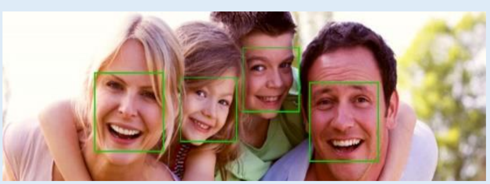
Machine Learning

- 1. Different ML methods
- *2. Data representation*
- 3. Data preprocessing

2. Data representation

Different types of inputs





Functions

Outputs

Different tasks

2. Data representation

A_1

A_2

A_3

A_4

A_5

A_6

- Feature: what is feature
- Apple = [diameter, color, shape, spots, place of production]
- Dimensionality: 5

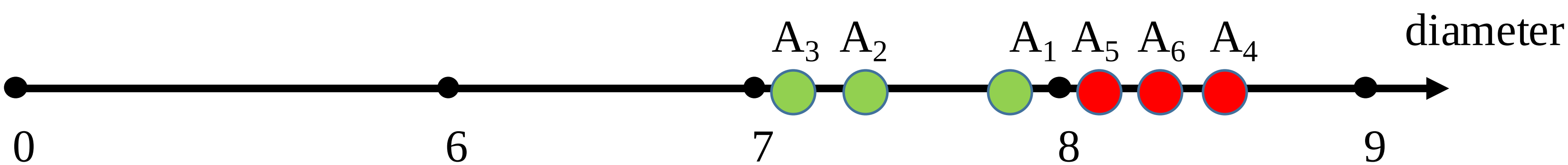
2. Data representation

Apple = [diameter, color, shape, spots, place of production]

A_1= [7.8]

A_2= [7.4]

A_3= [7.1]



A_4= [8.5]

A_5= [8.1]

A_6= [8.3]

2. Data representation

Apple = [diameter, color, shape, spots, place of production]

$A_1 = \begin{bmatrix} 7.8 \\ 0.2 \end{bmatrix}$

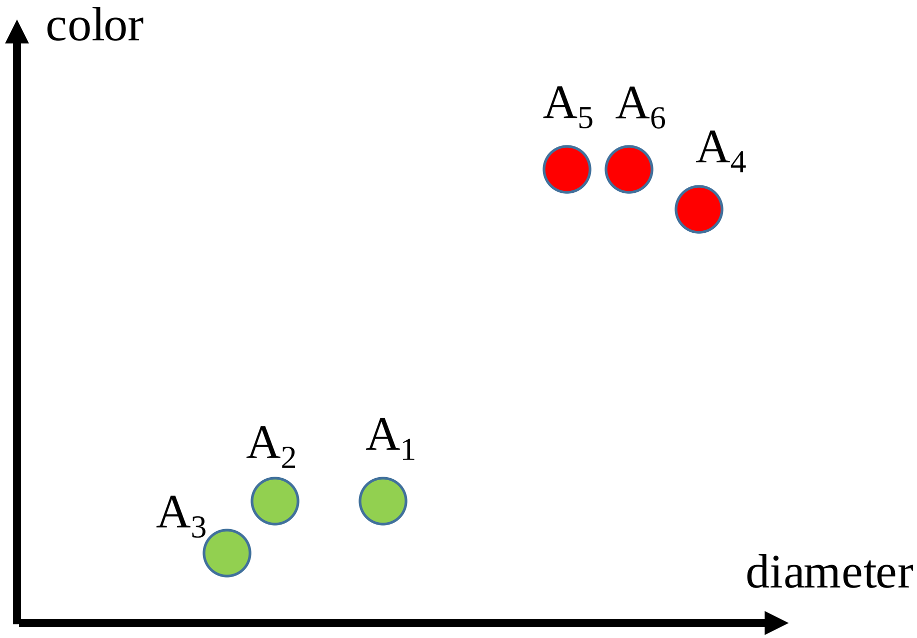
$A_2 = \begin{bmatrix} 7.4 \\ 0.2 \end{bmatrix}$

$A_3 = \begin{bmatrix} 7.1 \\ 0.1 \end{bmatrix}$

$A_4 = \begin{bmatrix} 8.5 \\ 0.7 \end{bmatrix}$

$A_5 = \begin{bmatrix} 8.1 \\ 0.8 \end{bmatrix}$

$A_6 = \begin{bmatrix} 8.3 \\ 0.8 \end{bmatrix}$



2. Data representation

Apple = [diameter, color, shape, spots, place of production]

$A_1 = \begin{bmatrix}7.8\\0.2\end{bmatrix}$

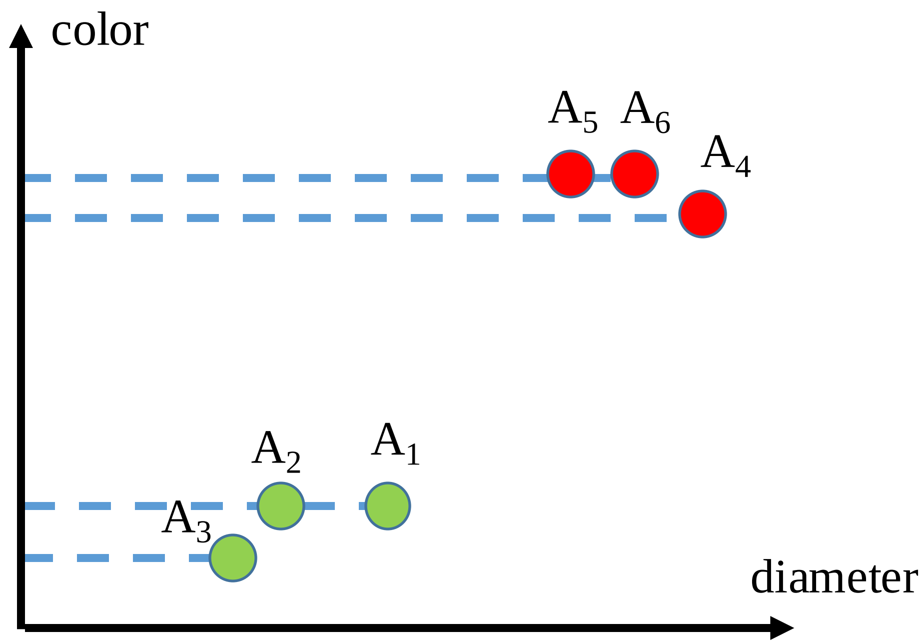
$A_2 = \begin{bmatrix}7.4\\0.2\end{bmatrix}$

$A_3 = \begin{bmatrix}7.1\\0.1\end{bmatrix}$

$A_4 = \begin{bmatrix}8.5\\0.7\end{bmatrix}$

$A_5 = \begin{bmatrix}8.1\\0.8\end{bmatrix}$

$A_6 = \begin{bmatrix}8.3\\0.8\end{bmatrix}$



2. Data representation

Apple = [diameter, color, shape, spots, place of production]

A_1= [0.2]

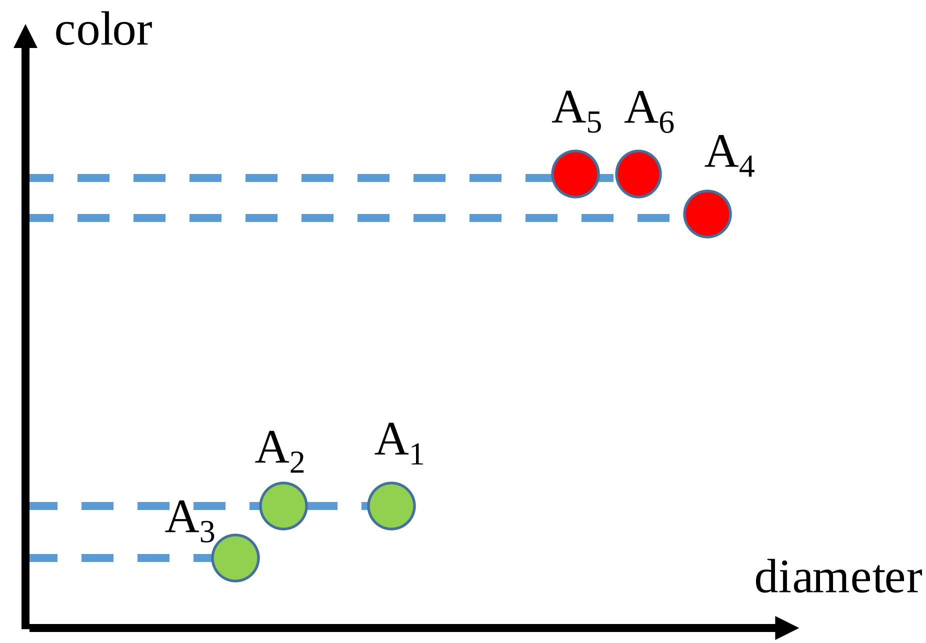
A_2= [0.2]

A_3=[0.1]

A_4= [0.7]

A_5= [0.8]

A_6= [0.8]



Dimensional reduction

2. Data representation

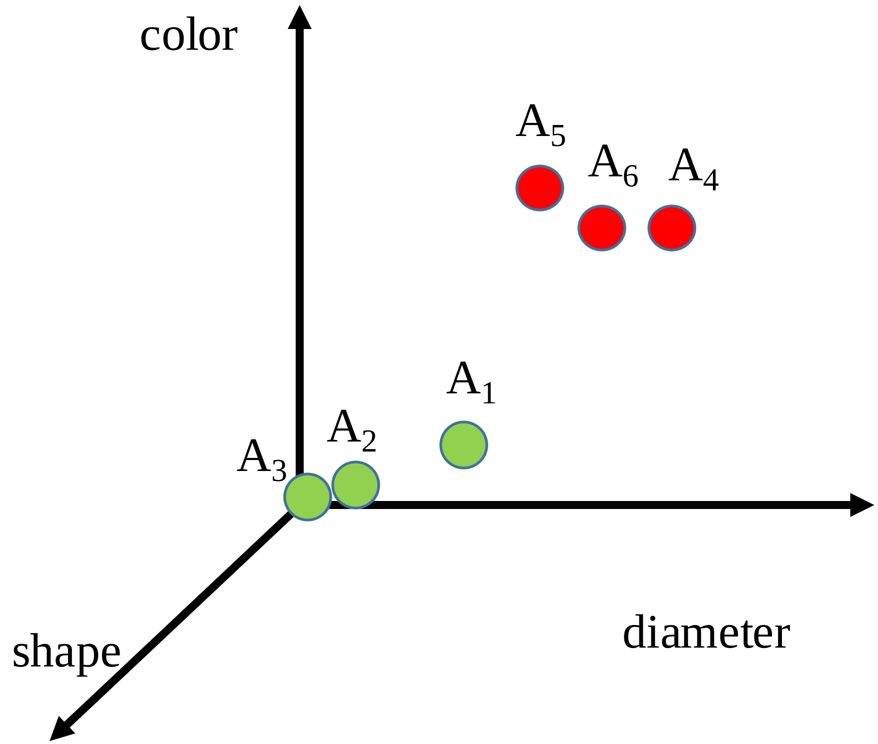
Apple = [diameter, color, shape, spots, place of production]



$A_1 = \begin{bmatrix} 7.8 \\ 0.2 \\ 0.6 \end{bmatrix}$ $A_2 = \begin{bmatrix} 7.4 \\ 0.2 \\ 0.7 \end{bmatrix}$ $A_3 = \begin{bmatrix} 7.1 \\ 0.1 \\ 0.6 \end{bmatrix}$



$A_4 = \begin{bmatrix} 8.5 \\ 0.7 \\ 0.7 \end{bmatrix}$ $A_5 = \begin{bmatrix} 8.1 \\ 0.8 \\ 0.7 \end{bmatrix}$ $A_6 = \begin{bmatrix} 8.3 \\ 0.8 \\ 0.8 \end{bmatrix}$



2. Data representation

Apple = [diameter, color, shape, spots, place of production]

$$A_1 = \begin{bmatrix} 7.8 \\ 0.2 \\ 0.6 \\ 1 \\ 1 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 7.4 \\ 0.2 \\ 0.7 \\ 0 \\ 1 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} 7.1 \\ 0.1 \\ 0.7 \\ 0 \\ 2 \end{bmatrix}$$

$$A_4 = \begin{bmatrix} 8.5 \\ 0.7 \\ 0.7 \\ 0 \\ 3 \end{bmatrix}$$

$$A_5 = \begin{bmatrix} 8.1 \\ 0.8 \\ 0.7 \\ 0 \\ 3 \end{bmatrix}$$

$$A_6 = \begin{bmatrix} 8.3 \\ 0.8 \\ 0.8 \\ 1 \\ 4 \end{bmatrix}$$

2. Data representation

Input: the values of the apples

$$\begin{bmatrix}7.8\\0.2\\0.6\\1\\1\end{bmatrix}\begin{bmatrix}8.1\\0.8\\0.7\\0\\3\end{bmatrix}\begin{bmatrix}7.4\\0.2\\0.7\\0\\1\end{bmatrix}\begin{bmatrix}7.1\\0.1\\0.7\\0\\2\end{bmatrix}\begin{bmatrix}8.5\\0.7\\0.7\\0\\3\end{bmatrix}\begin{bmatrix}8.3\\0.8\\0.8\\1\\4\end{bmatrix}$$

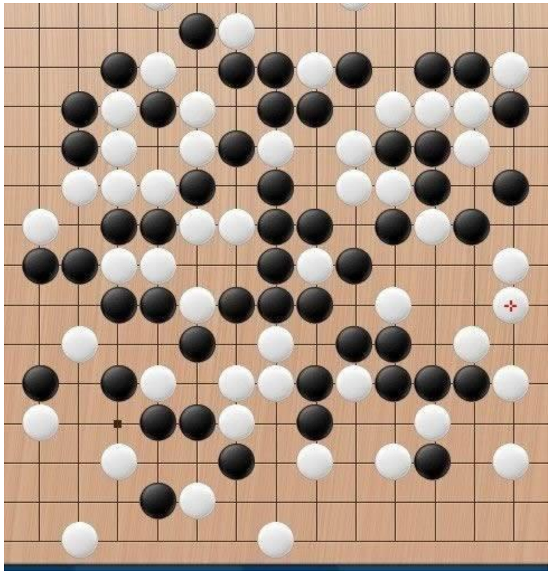
Output: the values of the apples

0 1

2. Data representation

Another example

Input: A certain state of the board



The state can be represented by a matrix.

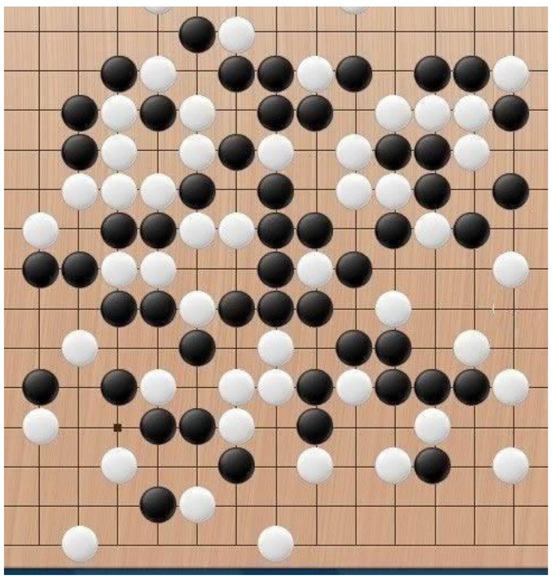
	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1
3	-1	0	0	-1	0	0	1	0	-1	0	0	1	-1
4	0	1	0	1	-1	0	0	-1	1	1	1	0	-1
5	0	1	-1	1	0	1	-1	1	0	0	1	-1	-1
6	1	-1	1	0	-1	0	-1	1	1	0	-1	0	-1
7	-1	0	0	1	1	0	0	-1	0	1	0	-1	-1
8	0	1	1	-1	-1	0	1	0	-1	-1	-1	1	-1
9	-1	0	0	1	0	0	0	-1	1	-1	-1	1	-1
10	1	-1	-1	0	-1	1	-1	0	0	-1	1	-1	-1
11	-1	0	1	-1	1	1	0	1	0	0	0	1	-1
12	-1	-1	0	0	1	-1	0	-1	-1	1	-1	-1	-1
13	-1	1	-1	-1	0	-1	1	-1	1	0	-1	1	-1
14	-1	-1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1
15	1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1

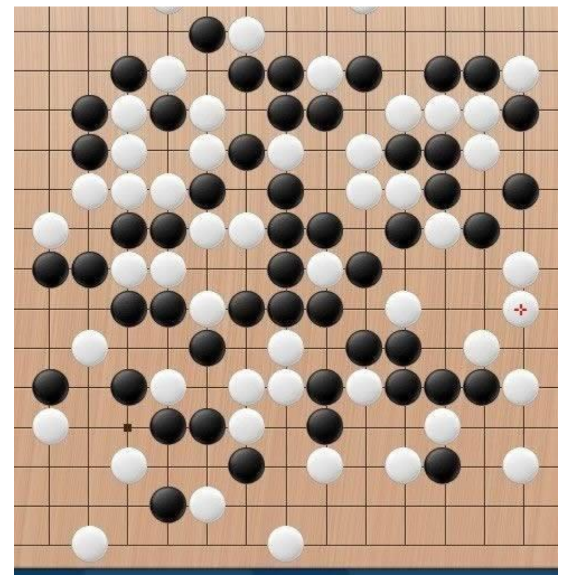
2. Data representation

- Another example

Input: A certain state of the board

Output: A new state after a move





2. Data representation

Another example

The input matrix.

	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1
3	-1	0	0	-1	0	0	1	0	-1	0	0	1	-1
4	0	1	0	1	-1	0	0	-1	1	1	1	0	-1
5	0	1	-1	1	0	1	-1	1	0	0	1	-1	-1
6	1	-1	1	0	-1	0	-1	1	1	0	-1	0	-1
7	-1	0	0	1	1	0	0	-1	0	1	0	-1	-1
8	0	1	1	-1	-1	0	1	0	-1	-1	-1	1	-1
9	-1	0	0	1	0	0	0	-1	1	-1	-1	1	-1
10	1	-1	-1	0	-1	1	-1	0	0	-1	1	-1	-1
11	-1	0	1	-1	1	1	0	1	0	0	0	1	-1
12	-1	-1	0	0	1	-1	0	-1	-1	1	-1	-1	-1
13	-1	1	-1	-1	0	-1	1	-1	1	0	-1	1	-1
14	-1	-1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1
15	1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1

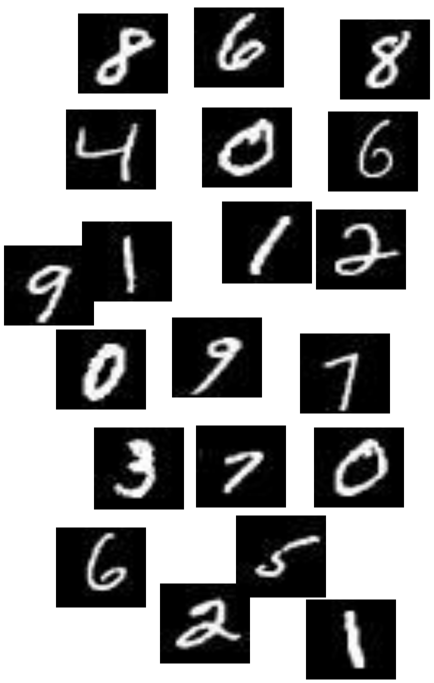
The output matrix.

	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1
3	-1	0	0	-1	0	0	1	0	-1	0	0	1	-1
4	0	1	0	1	-1	0	0	-1	1	1	1	0	-1
5	0	1	-1	1	0	1	-1	1	0	0	1	-1	-1
6	1	-1	1	0	-1	0	-1	1	1	0	-1	0	-1
7	-1	0	0	1	1	0	0	-1	0	1	0	-1	-1
8	0	1	1	-1	-1	0	1	0	-1	-1	-1	1	-1
9	-1	0	0	1	0	0	0	-1	1	-1	-1	1	-1
10	1	-1	-1	0	-1	1	-1	0	0	-1	1	-1	-1
11	-1	0	1	-1	1	1	0	1	0	0	0	1	-1
12	-1	-1	0	0	1	-1	0	-1	-1	1	-1	-1	-1
13	-1	1	-1	-1	0	-1	1	-1	1	0	-1	1	-1
14	-1	-1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1
15	1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1

2. Data representation

- 3rd example

Input: Images of size 28*28

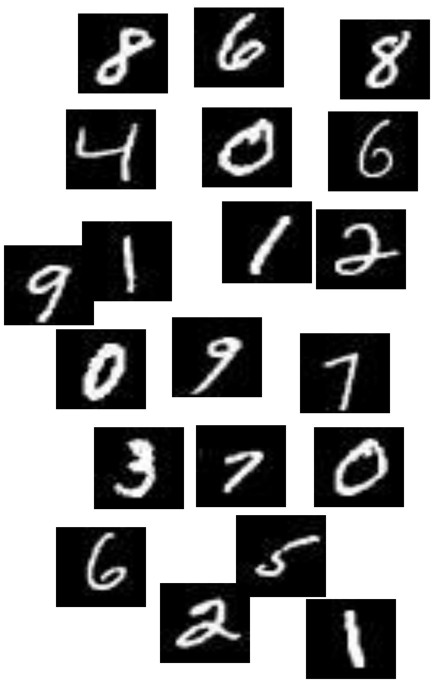


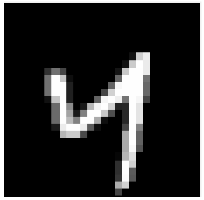
Output: Recognition results

8, 6, 8, 4, 0, 6…

2. Data representation

- 3rd example





Matrix of size 28*28

		5	6	7	8	9	10	11	12	13	14	15	16	17
6	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	0	0	0	49	112	69	0	0	0	0	0	0	0	
11	0	0	0	112	254	197	14	0	0	0	0	0	0	
12	0	0	0	112	254	254	32	0	0	0	0	0	99	
13	0	0	0	112	254	254	32	0	0	0	0	69	209	
14	0	0	0	17	195	254	32	0	0	0	100	245	254	
15	0	0	0	0	106	254	139	0	25	183	244	254	211	
16	0	0	0	0	106	254	162	25	128	254	254	200	78	
17	0	0	0	0	106	254	186	129	254	254	170	15	0	
18	0	0	0	0	27	236	254	254	236	91	15	0	0	
19	0	0	0	0	0	182	202	202	73	0	0	0	0	
20	0	0	0	0	0	0	0	0	0	0	0	0	0	
21	0	0	0	0	0	0	0	0	0	0	0	0	0	
22	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	0	0	0	0	0	0	0	0	0	0	0	0	0	
24	0	0	0	0	0	0	0	0	0	0	0	0	0	

Gray value: 0~255

2. Data representation

3rd example

The input matrix.

		5	6	7	8	9	10	11	12	13	14	15	16	17
6	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	0	0	0	49	112	69	0	0	0	0	0	0	0	
11	0	0	0	112	254	197	14	0	0	0	0	0	0	
12	0	0	0	112	254	254	32	0	0	0	0	0	99	
13	0	0	0	112	254	254	32	0	0	0	0	69	209	
14	0	0	0	17	195	254	32	0	0	0	100	245	254	
15	0	0	0	0	106	254	139	0	25	183	244	254	211	
16	0	0	0	0	106	254	162	25	128	254	254	200	78	
17	0	0	0	0	106	254	186	129	254	254	170	15	0	
18	0	0	0	0	27	236	254	254	236	91	15	0	0	
19	0	0	0	0	0	182	202	202	73	0	0	0	0	
20	0	0	0	0	0	0	0	0	0	0	0	0	0	
21	0	0	0	0	0	0	0	0	0	0	0	0	0	
22	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	0	0	0	0	0	0	0	0	0	0	0	0	0	
24	0	0	0	0	0	0	0	0	0	0	0	0	0	

The output **labels**.

8, 6, 8, 4, 0, 6…

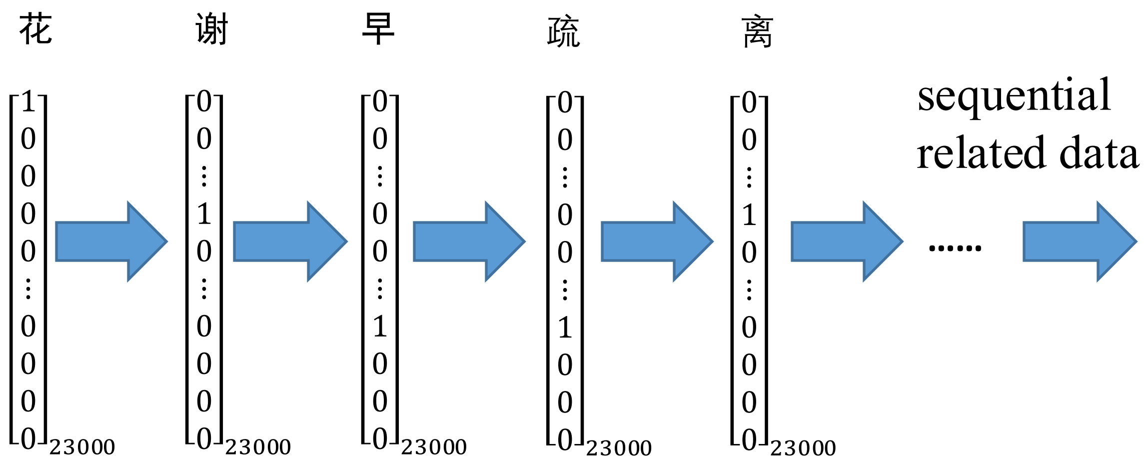
2. Data representation

4th example

How to generate a poem by computer?

卜算子·咏梅

花谢早疏篱，
几度陶潜里。
永日梅花昔底寒，
比向梅花妒。
荣悴幻非凡，
谓是娇芳伴。
后著金陵几日时，
中酒争先理。



Task: output the next word continually.

Machine Learning

- 1. Different ML methods
- 2. Data representation
- *3. Data preprocessing*

3. Data preprocessing

- Normalization
- Feature extraction
- Noise removal (image, speech, …)

3. Data preprocessing

Normalization

- Data normalization means transforming all variables in the data to a specific range.
- Two standard methods for normalization.
- 1. Normalizes the data so that they fall into a standard range

$$\mathbf{p}^n = 2(\mathbf{p} - \mathbf{p}^{min})./(\mathbf{p}^{max} - \mathbf{p}^{min}) - 1$$

- 2. Normalizes the data so that they have a specified mean and variance

$$\mathbf{p}^n = (\mathbf{p} - \mathbf{p}^{mean})./\mathbf{p}^{std}$$

3. Data preprocessing

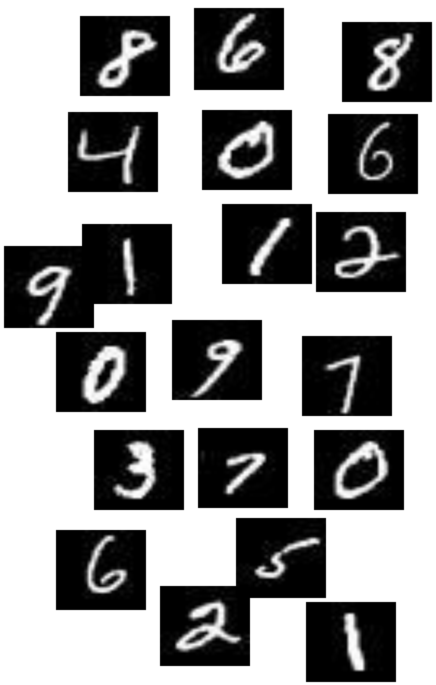
Feature extraction

- Feature extraction is the **transformation** of the original data (using all variables/features) to a dataset with a reduced number of features.
- In feature extraction, all available features are used and the data are transformed (using a linear or nonlinear transformation) to a reduced dimension space. Thus, the aim is to replace the original features by a smaller set of underlying features.

3. Data preprocessing

Feature extraction

- How to select and extract features?
- Depends on the problem.



- Gray value: 0~255

		5	6	7	8	9	10	11	12	13	14	15	16	17
6	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	0	0	0	49	112	69	0	0	0	0	0	0	0	
11	0	0	0	112	254	197	14	0	0	0	0	0	0	
12	0	0	0	112	254	254	32	0	0	0	0	0	99	
13	0	0	0	112	254	254	32	0	0	0	0	69	209	
14	0	0	0	17	195	254	32	0	0	0	100	245	254	
15	0	0	0	0	106	254	139	0	25	183	244	254	211	
16	0	0	0	0	106	254	162	25	128	254	254	200	78	
17	0	0	0	0	106	254	186	129	254	254	170	15	0	
18	0	0	0	0	27	236	254	254	236	91	15	0	0	
19	0	0	0	0	0	182	202	202	73	0	0	0	0	
20	0	0	0	0	0	0	0	0	0	0	0	0	0	
21	0	0	0	0	0	0	0	0	0	0	0	0	0	
22	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	0	0	0	0	0	0	0	0	0	0	0	0	0	
24	0	0	0	0	0	0	0	0	0	0	0	0	0	

- Image features: edge feature map
- Extract features by some algorithms such as PCA, ICA, DNN…
- ……

3. Data preprocessing

Feature extraction

- Linear feature extraction

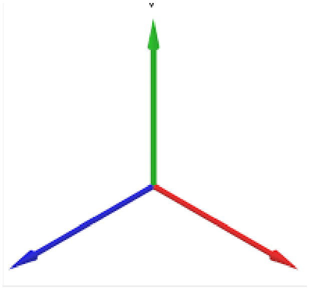
➢ Given the original d-dimension feature space $X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m) \in \mathbb{R}^{d\times m}$

➢ Get the reduced d'-dimension feature space $Z = (\boldsymbol{z}_1, \boldsymbol{z}_2, \dots, \boldsymbol{z}_m) \in \mathbb{R}^{d'\times m}$ after transformation ($d' < d$)

➢ Transformation process:

$$Z = \boldsymbol{W}^T X$$

Where $\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_{d'}) \in \mathbb{R}^{d\times d'}$ is the transformation matrix, $\boldsymbol{w}_i \in \mathbb{R}^{d\times 1}$, and $Z \in \mathbb{R}^{d'\times m}$ is the coordinate expression of sample X in low dimension space.



3. Data preprocessing

- Noise removal

before

after

Conclusion

- Different ML ethods
  - Brief introduction to supervised learning, unsupervised learning and reinforcement learning
- Data representation
- Data preprocessing
  - Normalization
  - Feature extraction
  - Noise removal (image, speech, …)